

# **$\mathcal{J}$ FIT: a framework to obtain combined experimental results through joint fits**

E. Ben-Haim<sup>1</sup>, R. Brun<sup>2</sup>, B. Echenard<sup>3</sup>, T.E. Latham<sup>4</sup>

<sup>1</sup>*Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE),  
IN2P3/CNRS, Université Pierre et Marie Curie-Paris 6, Université Denis  
Diderot-Paris 7*

<sup>2</sup>*European Organization for Nuclear Research (CERN), Geneva, Switzerland*

<sup>3</sup>*California Institute of Technology, Pasadena, California 91125, USA*

<sup>4</sup>*Department of Physics, University of Warwick, Coventry CV4 7AL, United  
Kingdom*

## **Abstract**

A master-worker architecture is presented for obtaining combined experimental results through joint fits of datasets from several experiments, ensuring that correlations are correctly taken into account and resulting in a better determination of nuisance parameters. The  $\mathcal{J}$ FIT framework allows such joint fits to be performed keeping the data separated, in its original format, and using independent fitting environments. We present a C++ implementation of such a framework based on the ROOT package, and demonstrate its functionalities with concrete examples.

## **1 Introduction**

In the vast majority of cases, high energy physics experiments obtain measurements by performing a maximum-likelihood fit to their data to extract observables of interest. Combining results of different experiments is routinely performed, using different well-established methods and dedicated tools. In the simplest approach, which is often used, measured observables are assumed to be normally distributed and standard statistical prescriptions are readily applied. However, there are situations in which the procedure is challenging even with the simple normal-distribution hypothesis, in particular for measurements involving a large number of parameters (including nuisance parameters). The size of the covariance matrices, when available, makes the procedure tedious and prone to errors. When non-Gaussian uncertainties are taken into account the combination procedure becomes much more complicated, as working with the actual likelihood functions is generally needed. This is indispensable in many complex measurements, in cases with small sample sizes and combination of upper limits. Sometimes only a partial likelihood function is considered, which is the projection of the full function on several parameters of interest, assuming that they are uncorrelated with the other parameters.

Ideally, combining measurements should be done by fitting simultaneously the different data samples. The straightforward solution involves collecting data in a common format, and has been implemented, among others, in the RooFit package [1]. This solution could however be inefficient if dedicated fitting frameworks

have already been developed, especially for complex fit models. It may also raise issues with data access policies of each experiment. An alternative approach consists of performing a joint fit while keeping the data separated. This can be achieved using a master-worker architecture, in which the master drives the fit by combining the values of the likelihood functions returned by several workers, each of which is specific to an experiment and accesses its own dataset. This solution presents several advantages:

- the data do not need to be rewritten in any external format, and can be readily used;
- experimental collaborations can keep private their data and procedures;
- any fitting algorithm can be readily incorporated in this scheme;
- all correlations are transparently taken into account;
- correlated and uncorrelated systematic uncertainties are easily included;
- common nuisance parameters are also determined jointly, allowing further improvement in precision.

In this paper we present **JFIT**: an implementation of this idea within the ROOT framework [2]. In Sec. 2, we detail the **JFIT** joint fitting method (a general brief review of the maximum-likelihood technique is given in Appendix A). Two examples of application of the method in the domain of high energy physics are briefly described in Sec. 3 (with further details given in Appendix B): a measurement of properties of a resonance, and an amplitude analysis of charmless  $B$ -meson decays with a three-body final state.

## 2 Joint fits and the **JFIT** method

The formalism to combine several measurements by performing a joint maximum-likelihood fit of different datasets (e.g., from different experiments) is well known: the likelihood  $\mathcal{L}$  to maximise is given by the product of likelihoods from the different datasets<sup>1</sup>. In the case of two datasets  $A$  and  $B$  (the generalisation to a larger number of datasets is straightforward), with individual likelihoods  $\mathcal{L}_A$  and  $\mathcal{L}_B$ , the combined likelihood is written as:

$$\mathcal{L}(\theta, \theta_A, \theta_B) = \prod_{x \in x_A} \mathcal{L}_A(\theta, \theta_A) \prod_{x \in x_B} \mathcal{L}_B(\theta, \theta_B), \quad (1)$$

where  $x$  is a random variable (or a set of several random variables) with corresponding ensembles of independent observations, designated by  $x_A$  and  $x_B$ , in the two datasets;  $\theta$  denotes the parameters of interest that are shared by the two datasets, and have to be simultaneously extracted from both (common parameters);  $\theta_A$  and  $\theta_B$  are parameters that are specific for  $A$  and  $B$  (specific parameters). A simple

---

<sup>1</sup> A brief review of the maximum-likelihood estimation method is given in Appendix A.

example would be a one-dimensional fit to extract a common parameter of interest, such as the mass of a resonance, from two datasets in which the signal and background distributions are modelled differently, using specific parameters.

In the **JFIT** approach, the master drives the fit by sending sets of parameters to the experiment-specific workers. Each worker returns the value of the likelihood function to the master, which in turn updates the parameters using an optimisation routine, such as the MIGRAD algorithm of the MINUIT optimization package [3] widely used in the field of high energy physics. The procedure is repeated until the fit has converged. This architecture keeps the individual likelihood functions  $\mathcal{L}_A$  and  $\mathcal{L}_B$  separate, and allows a flexible treatment of the specific parameters. The latter can either be controlled by the master, or declared only within the corresponding worker. In the second case, the workers perform an additional minimization with respect to their specific parameters at each step of the procedure, keeping the common parameters fixed. The values of the minimized functions are then returned to the master.

A specific open source implementation of **JFIT** in C++ has been developed within the Laura++ Dalitz-plot analysis package [4], using the ROOT framework [2]. The structure of our implementation is based on the following steps:

- Initialization: both master and workers initialize their internal structure to either drive the fit (master) or calculate the likelihood given a set of parameters (workers).
- Synchronization: the master-worker communications are handled using several classes in the ROOT framework<sup>2</sup>: **TMonitor**, **TServerSocket** and **TSocket**. The master starts a server, and waits for the worker nodes to connect. The connections can be secured via ssh if required. After successfully connecting, each worker awaits instructions from the master. There then follows an exchange of information (using the **TMessage** class) regarding the parameters known to each worker and their initial values.
- Minimisation: the master starts the fitting procedure by sending a set of parameter values to the workers (again via the **TMessage** class) and waits for their replies. Upon receipt of the parameters, each worker calculates the value of its corresponding likelihood function<sup>3</sup> and sends the result back to the master. The worker results are then combined by the master and given to the optimiser, which updates the fit parameters and the master then sends a new request to the workers. This step is repeated until convergence is achieved. The uncertainties on the parameters are evaluated by a similar procedure.
- Termination: the master returns the results of the fit and terminates the connections with the workers.

---

<sup>2</sup> This implementation is based on the ROOT macros:  
<http://root.cern.ch/root/html/tutorials/net/authserv.C.html>  
<http://root.cern.ch/root/html/tutorials/net/authclient.C.html>

<sup>3</sup> If specific parameters are declared inside a worker, the likelihood is evaluated by performing a secondary minimization over these parameters (keeping the common parameters fixed).

The master routines are implemented within the `LauSimFitMaster` class in `Laura++` [5], while the communication methods of the workers are implemented in the `LauSimFitSlave` base class [6]. A large fraction of code developed by each experiment to perform dedicated fits can be readily reused in this scheme; it requires only the addition of a layer that inherits from `LauSimFitSlave`, which should call the appropriate parts of the pre-existing code to perform the actions required at each stage of the fit (e.g., calculating the likelihood value). Internal changes in one experiment (e.g. new data format or additional specific parameters) are completely transparent. It is worth mentioning that this framework can also be used to perform simultaneous fits to different datasets within an experiment, for example to account for different trigger categories, as in Ref. [7]. It can also be used to parallelise fits across multiple cores by splitting the dataset into smaller samples, each of which is processed by a dedicated worker.

The overhead introduced by the  $\mathbf{JFIT}$  framework itself is small. For instance, in the second example in Sec. 3 (that is detailed in Appendix B.2), the time for each individual fit to be performed was  $33 \pm 6$  seconds, while the joint fits using  $\mathbf{JFIT}$  took  $35 \pm 5$  seconds.

### 3 Benefits of joint fitting as demonstrated by example applications of $\mathbf{JFIT}$

Joint fitting has many advantages compared to other methods to combine results from two experiments. In this section we briefly review the benefits obtained from the application of the  $\mathbf{JFIT}$  framework compared to naïve averaging of individual results, using two examples from the domain of high energy physics. The examples are explained in detail in Appendix B.

As a first example (Appendix B.1) we examine the problem of combining the results of two experiments that found a resonance in an invariant mass spectrum. The parameters of interest, the mass of the resonance and its branching fraction to the observed final state, are obtained from a one-dimensional maximum-likelihood fit to the distribution of invariant mass of the final-state particles, including signal and background components. This example demonstrates the following advantages of performing a joint fit to combine results from two experiments:

- the estimation of statistical uncertainties is more reliable;
- the estimation of systematic uncertainties is more straightforward and also more reliable.

In particular, systematic uncertainties originating from external sources, such as those related to a measured branching fraction of a background, are taken into account in a simple and coherent way. While these are usually assumed to be fully correlated in a naïve combination of results, in reality different patterns of migration between event species in the two experiments may alter this behaviour.

The second example (Appendix B.2) considers an amplitude analysis of a three-body  $B$ -meson decay, where the parameters of interest are those that characterise the

violation of the Charge conjugation-Parity ( $CP$ ) symmetry in the amplitudes of the contributing intermediate resonances. This type of analysis usually involves a large number of fit parameters with non-trivial correlations. This example demonstrates the following additional advantages of performing a joint fit under such circumstances:

- the central values of results are less likely to be biased since all the correlations between the parameters of interest and the nuisance parameters are included;
- common nuisance parameters can be better constrained, which leads to improved precision on the parameters of interest.

In addition, it is noted that there are other positive side effects from obliging collaborations to coordinate their models and conventions prior to performing their analyses.

## 4 Conclusions

In this paper, we have presented **JFIT**: a framework for obtaining combined experimental results through joint fits of datasets from several experiments. Using a master-server architecture, the tool allows such fits to be performed keeping the data separated and using independent fitting frameworks. It simplifies the process with respect to data access policies and allows existing code to be used with minimal changes, thus saving resources. A C++ implementation of the tool can be found in the Laura++ Dalitz-plot analysis package [4], using the network communication classes from the ROOT framework (`TMessage`, `TMonitor`, `TServerSocket` and `TSocket`).

The usefulness of the joint fitting approach has been illustrated by two example applications of **JFIT**: one from the domain of resonance searches, and the other from amplitude analyses in flavour physics. Correctly accounting for correlations between parameters in likelihood functions in different experiments, which is an intrinsic property of joint fits, results in improved results, and more reliable uncertainties. Thus, these fits provide a means to better exploit data from multiple experiments.

## Acknowledgments

The work presented in this paper was started in the purpose of performing joint analyses between the *BABAR* and Belle collaborations. During his mandate as the *BABAR* spokesperson, François Le Diberder initiated this project and generated interest in it among the *BABAR* and Belle collaborations, and the authors of ROOT. We would like to thank him for these actions and for fruitful discussions concerning this paper. We would also like to thank Roger Barlow for his extremely valuable advice and comments regarding the statistical aspects and examples. This work is supported by the Science and Technology Facilities Council (United Kingdom), the European Research Council under FP7, and the US Department of Energy under grants DE-FG02-92ER40701 and DE-SC0011925.

# Appendix

## A Maximum-likelihood estimation

Maximum-likelihood estimation is a widely-used method of fitting parameters of a model to some data and providing an estimate of their uncertainties. Here we briefly review this technique; more details can be found in statistics textbooks, and, for example, in Ref. [8].

Consider a set of  $N$  independent observations  $x_1, \dots, x_N$  of a random variable  $x$  following a probability density function (PDF) modelled by  $\mathcal{P}(x, \theta)$ , where  $\theta$  denotes the unknown model parameter(s) that should be estimated. Both  $x$  and  $\theta$  can be multidimensional. The likelihood function  $\mathcal{L}$  is defined as:

$$\mathcal{L}(\theta) = \prod_{i=1}^N \mathcal{P}(x_i, \theta). \quad (2)$$

Since the  $N$  observations  $x_i$  are given, the likelihood is only a function of  $\theta$ . The resulting estimate  $\hat{\theta}$  of the parameter  $\theta$  is defined as the value that maximises the likelihood, i.e.

$$\hat{\theta} = \max_{\theta} \mathcal{L}(\theta). \quad (3)$$

It is often convenient to minimise the negative logarithm of the likelihood (NLL) instead of maximising the likelihood, which yields the same value of  $\hat{\theta}$ . If the number of observations is random, an additional term is usually included in the likelihood to form the so-called extended likelihood:

$$\mathcal{L}_{\text{ext}}(\theta) = \frac{e^{-n} n^N}{N!} \prod_{i=1}^N \mathcal{P}(x_i, \theta), \quad (4)$$

where  $n$  denotes the expected number of events. Obtaining combined measurements from several datasets is achieved by maximising the product of likelihoods from the different datasets, as discussed in Sec. 2.

The minimisation of the NLL can be, and usually is, done numerically. Among the packages available to perform this task, MINUIT [3] is one of the most popular in the field of high energy physics. Its default algorithm, MIGRAD, is based on a variable-metric method that computes the value of the function and its gradient at each step of the procedure. If the functional form of the derivative is not, or cannot be supplied by the user, the gradient is evaluated by finite differences. The minimisation stops when the difference of the value of the function between two successive steps reaches a specific threshold. Parabolic uncertainties on the parameters are estimated by inverting the matrix of second derivatives evaluated at the minimum of the function. In the case of non-Gaussian likelihood, the uncertainties are estimated by an algorithm that scans the likelihood for each parameter separately, minimising the likelihood each time with respect to the remaining parameters.

## B Details of the example applications of $\mathcal{J}_{\text{FIT}}$

### B.1 Resonance mass and branching-fraction measurement

As a first example, we examine the problem of combining the results of two experiments that found a resonance in an invariant mass spectrum, at a mass of  $126 \text{ GeV}/c^2$ . The parameters of interest, the mass of the resonance,  $m_{\text{Res}}$ , and its branching fraction, BF, to the observed final state, are obtained from a one-dimensional maximum-likelihood fit to the distribution of invariant mass,  $m$ , of the final-state particles.

Samples corresponding to the datasets of experiments 1 and 2 are generated in the invariant-mass range of  $[100, 160] \text{ GeV}/c^2$  according to a model containing three events species: signal, peaking background and combinatorial background. Signal events are generated from a sum of two Crystal Ball (CB) functions [9]: core and tail, with different peak positions. Peaking and combinatorial background events are generated from a Gaussian and a fourth order Chebyshev polynomial function, respectively. This model is roughly inspired by the search for the Standard Model Higgs boson from ATLAS [10] and CMS [11], although, with rather different event yields and signal to background ratios. As the underlying physics is assumed to be unique, the same model is used to generate events for both experiments. To emulate in a simple way different overall efficiencies and event-selection strategies, the datasets corresponding to the two experiments have significantly different numbers of signal and background events. Five hundred pairs of datasets are generated, where the numbers of events of each species are Poisson-distributed, using PDFs that are summarised in Tab. 1. The invariant mass distribution of experiment 1 is shown in Fig. 1.

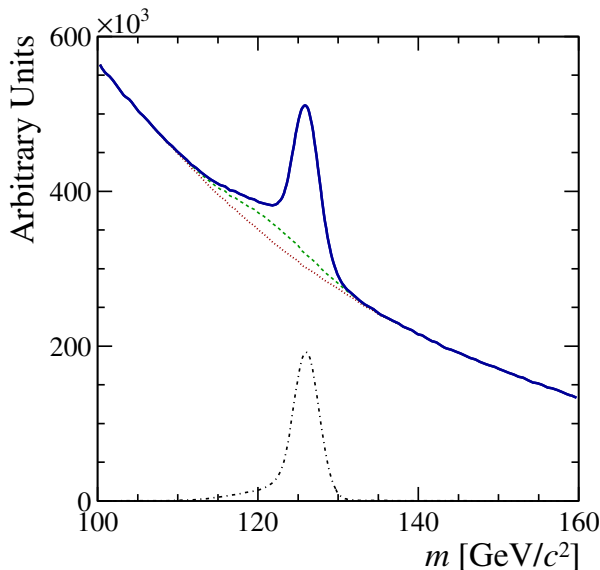


Figure 1: The invariant-mass distribution of a sample generated for experiment 1. The solid (blue) curve shows the full sample. The dash-dotted (black) curve corresponds to the distribution of signal events, while the dotted (red) and the dashed (green) curves show the distributions for the combinatorial and total background, respectively.

Table 1: Summary of the model used to generate events. The average number of events of each species generated per experiment is denoted  $N$ . The peak positions and Gaussian widths of the CB and Gaussian functions are denoted  $\mu$  and  $\sigma$ , respectively, with the superscript core or tail, and with the subscript CB, as appropriate. The CB tail parameters are  $\alpha$  and  $n$ . The coefficient of the  $i^{\text{th}}$  power term in the polynomial is denoted  $c_i$ .

Event species	Function	Parameter	Value
Signal	Sum of two CB functions (core and tail)	$N$ (experiment 1)	3000
		$N$ (experiment 2)	10000
		$\mu_{\text{CB}}^{\text{core}}$	126 GeV/ $c^2$
		$\sigma_{\text{CB}}^{\text{core}}$	1.6 GeV/ $c^2$
		$\alpha^{\text{core}}$	1.5
		$n^{\text{core}}$	3.0
		Fraction of tail	10%
		$\mu_{\text{CB}}^{\text{tail}}$	120 GeV/ $c^2$
		$\sigma_{\text{CB}}^{\text{tail}}$	4.0 GeV/ $c^2$
		$\alpha^{\text{tail}}$	-1.0
		$n^{\text{tail}}$	4.0
Peaking background	Gaussian	$N$ (experiment 1)	1000
		$N$ (experiment 2)	7500
		$\mu$	122 GeV/ $c^2$
		$\sigma$	5.0 GeV/ $c^2$
Combinatorial background	4 <sup>th</sup> order Chebyshev polynomial	$N$ (experiment 1)	60000
		$N$ (experiment 2)	500000
		$c_1$	-0.682
		$c_2$	0.122 [GeV/ $c^2$ ] <sup>-1</sup>
		$c_3$	-0.013 [GeV/ $c^2$ ] <sup>-2</sup>
		$c_4$	-0.003 [GeV/ $c^2$ ] <sup>-3</sup>



The maximum-likelihood fits are performed, in the same  $[100, 160]$   $\text{GeV}/c^2$  invariant-mass range. For simplicity, the same model is used for the two experiments. However, the fitted model differs slightly from that used to generate events in order to account for the lack of knowledge of the underlying physics. The invariant mass of the signal is modelled by the sum of a Crystal Ball function, describing the core distribution, and a Gaussian describing the tails. Backgrounds are modelled by the same functional forms used to generate events. The signal peak position and all the combinatorial background parameters are varied in the fit, as well as the yields of the three events species. All the other parameters are fixed. The PDFs used in the fits are summarised in Tab. 2.

Table 2: Summary of the model used to fit events. Parameter notations are the same as in Tab. 1. Values of fixed parameters are given in the “Value” column; “Gen.” means that the fixed value is identical to that used for generation. The parameters  $\mu^{\text{tail}}$  and  $\mu_{\text{CB}}^{\text{core}}$  are constrained to take the same value.

Event species	Function	Parameter	Value
Signal	Sum of a core CB function and a Gaussian tail	$N$	varied
		$\mu_{\text{CB}}^{\text{core}}$	varied
		$\sigma_{\text{CB}}^{\text{core}}$	$1.5 \text{ GeV}/c^2$
		$\alpha^{\text{core}}, n^{\text{core}}$	Gen.
		Fraction of tail	Gen.
		$\mu^{\text{tail}}$	$= \mu_{\text{CB}}^{\text{core}}$
		$\sigma^{\text{tail}}$	$3.8 \text{ GeV}/c^2$
Peaking background	Gaussian	$N$	varied
		$\mu$	$122 \text{ GeV}/c^2$
		$\sigma$	$5 \text{ GeV}/c^2$
Combinatorial background	4 <sup>th</sup> order Chebyshev polynomial	$N$	varied
		$c_1, c_2, c_3, c_4$	varied

After fitting each of the individual samples, we obtain combined results for the mass of the resonance and the branching fraction by two different methods:

1. a naïve average of the individual results, taking into account the parabolic uncertainties of the fits to the individual samples;
2. a joint fit to the two samples in the **JFIT** framework, following the method proposed in this paper, using the full likelihood function.

As the likelihood functions are fairly Gaussian, asymmetric uncertainties have not been considered for the naïve averaging. The typical statistical uncertainty obtained by the two methods is 0.0027 for the branching fraction and  $0.054 \text{ GeV}/c^2$  for the mass of the resonance. Moreover, no bias has been observed in the extraction of the mass of the resonance by the two methods, while they both show a negative bias in the extraction of the branching fraction, as expected from the differences between generation and fit models. The results of the two methods for the resonance

branching fraction are compared in Fig. 2. In this particular case, the bias induced by the two methods is similar and thus the central values are in good agreement. There is a small effect on the statistical uncertainty, which is very slightly overestimated in naïve averages. This behaviour may differ, in size and direction, in other cases. For instance, the statistical uncertainty on the mass of the resonance was, conversely, found to be underestimated in naïve averages. However, as the joint fits correctly take into account all correlations, they provide more reliable results.

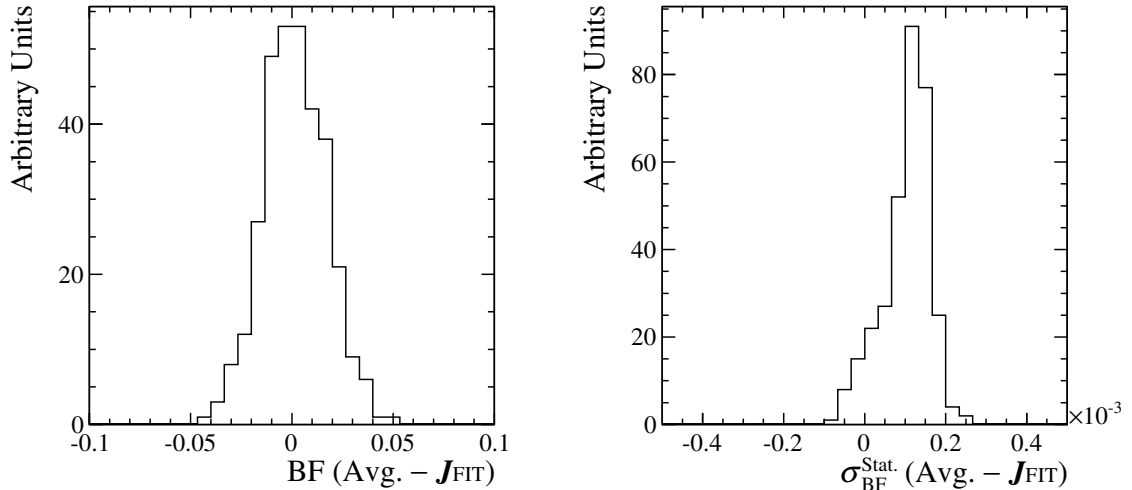


Figure 2: The distribution of the difference between the results of combinations performed by naïve averaging and  $\mathbf{J}_{\text{FIT}}$ , for the (left) central values of the resonance branching fraction, and the (right) corresponding statistical uncertainties. The former shows that there is no difference in the bias induced by the two methods, and the latter shows that the statistical uncertainty (typically 0.0027) is overestimated by 5% in naïve averages.

We expect larger differences between the two combination methods to arise when systematic uncertainties are evaluated. Parameters of the peaking background are considered to be badly known and, as such, to be sources of systematic uncertainties. To evaluate the effect from the width of the distribution, the corresponding PDF parameter is fixed to values  $2 \text{ GeV}/c^2$  above and below its nominal value, namely  $3.0 \text{ GeV}/c^2$  and  $7.0 \text{ GeV}/c^2$ . The variations of the resulting resonance mass and branching fraction are considered as systematic uncertainties. The same procedure is applied in individual and joint fits. Figure 3 shows a comparison between the systematic uncertainties obtained in naïve averages, where the individual systematic effects are considered to be 100% correlated, and in  $\mathbf{J}_{\text{FIT}}$ s. Differences are due to the fact that correlations are correctly taken into account in joint fitting. The comparison shows clearly that the effect on systematic uncertainties can be large (in some cases more than twice the statistical uncertainty or 10% of the systematic uncertainty), and they may be either underestimated or overestimated by averaging the results.

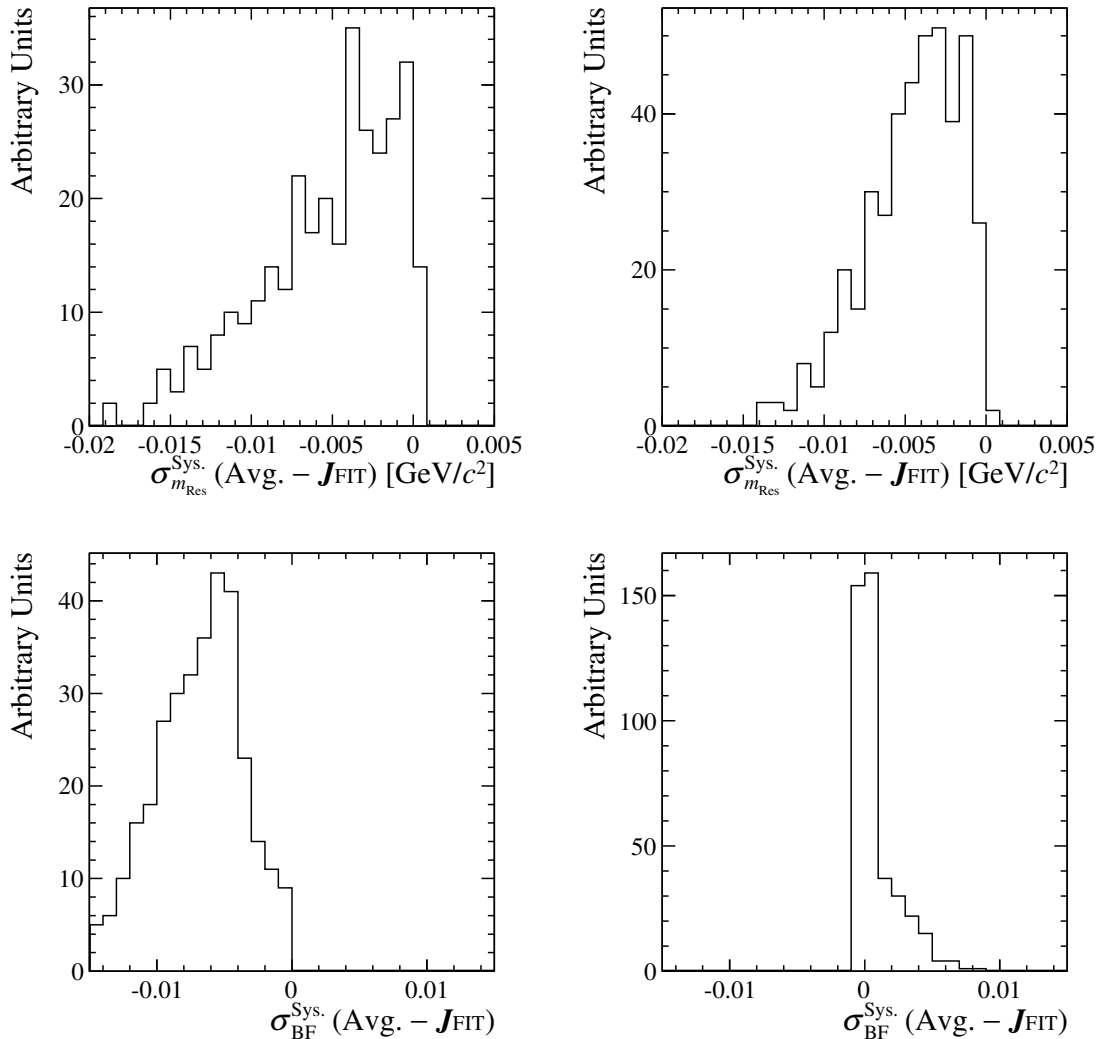


Figure 3: The difference between the systematic uncertainty values obtained in naïve averages and in  $\mathbf{J}_{\text{FIT}}$ s. The systematic uncertainty considered is that arising from the width of the peaking background component. The naïve averages assume that the systematic effects are 100% correlated between the two experiments. The effect on both (top)  $m_{\text{Res}}$  and (bottom) the branching fraction, is shown for a peaking background Gaussian width of (left)  $3.0 \text{ GeV}/c^2$  and (right)  $7.0 \text{ GeV}/c^2$ . In most of the cases the uncertainty is underestimated by naïve averaging. The exception is the uncertainty on the branching fraction obtained by increasing the width to  $7.0 \text{ GeV}/c^2$  (bottom right plot). The typical statistical uncertainty on the resonance mass is  $0.054 \text{ GeV}/c^2$ , while that of the branching fraction is  $0.0027$ .

## B.2 Amplitude analysis of a three-body $B$ -meson decay

A second example is from the domain of flavour physics. It involves a larger number of parameters in the fit, and illustrates other advantages of joint fitting. For simplicity, we consider signal events only.

A key part of the physics programme of the  $B$ -factories, *BABAR* and *Belle*, con-

sisted of amplitude (Dalitz-plot) analyses of 3-body  $B$ -meson decays [12]. More recently such studies are performed by the LHCb experiment (see, for example, Refs. [13, 14]), and, in the near future, will be also undertaken by Belle-II. These analyses provide measurements of CKM angles and access to observables sensitive to physics beyond the Standard Model as well as information on the resonant structure of decays. At the same time, such analyses have a strong dependence on model assumptions. In general, these analyses are limited by the sample size and therefore working with a larger, common dataset could be a fruitful approach. Besides the benefit of grouping the expertise of the different collaborations, a joint analysis has many advantages compared to a simple combination of results from separate analyses. Firstly, a joint analysis provides a more powerful determination of which components should be in the signal model, and allows setting better limits on minor components. It is important to notice that minor, poorly determined signal components are a major source of the so called model uncertainty that is often a large systematic effect in Dalitz-plot analyses. Secondly, the fact that different collaborations often use different parameterisations of resonant decay modes in the signal model (e.g., resonance lineshapes and phase conventions) can lead to difficulties in comparing their results. In some cases a direct comparison of such results can be less meaningful; they are less useful for the community, and averaging them becomes non trivial. The coordination of signal models, which is a *sine qua non* for a joint fit, is therefore beneficial. These two advantages of joint fits are not explicitly illustrated in this work.

To exemplify direct advantages of joint fitting, we use the result of the *BABAR* Dalitz-plot analysis of  $B^\pm \rightarrow K^\pm \pi^\mp \pi^\pm$  decays [15]. This analysis provided  $CP$ -averaged branching fractions and direct  $CP$  asymmetries for intermediate resonant and non-resonant contributions. It reported evidence for direct  $CP$  violation in the decay  $B^\pm \rightarrow \rho^0(770)K^\pm$ , with a  $CP$ -violation parameter  $A_{CP} = (44 \pm 10 \pm 4_{-13}^{+5})\%$ , where the first quoted uncertainty is statistical, the second is systematic, and the third is the model uncertainty mentioned above. The Belle collaboration also reported evidence of direct  $CP$  violation in the same decay mode [16] with a similar significance. This gives a strong motivation to obtain a combined result.

In the *BABAR* analysis, the contributions of the different intermediate states in the decay were obtained from a maximum-likelihood fit of the distribution of events in the Dalitz plot formed from the two variables  $m_{K\pi}^2 \equiv m_{K^\pm \pi^\mp}^2$  and  $m_{\pi\pi}^2 \equiv m_{\pi^\pm \pi^\mp}^2$ . As in many other Dalitz-plot analyses, the total signal amplitudes  $A$  and  $\bar{A}$  for  $B^+$  and  $B^-$  decays, respectively, were given in the isobar formalism, by

$$A = A(m_{K\pi}^2, m_{\pi\pi}^2) = \sum_j c_j F_j(m_{K\pi}^2, m_{\pi\pi}^2) \quad (5)$$

$$\bar{A} = \bar{A}(m_{K\pi}^2, m_{\pi\pi}^2) = \sum_j \bar{c}_j \bar{F}_j(m_{K\pi}^2, m_{\pi\pi}^2), \quad (6)$$

where  $j$  is a given intermediate decay mode. The distributions  $F_j \equiv \bar{F}_j$  are the lineshapes (e.g., Breit–Wigner functions) describing the dynamics of the decay amplitudes, and the complex coefficients  $c_j$  and  $\bar{c}_j$  contain all the weak-phase dependence and are measured relative to one of the contributing channels. They were

parameterised as

$$\begin{aligned} c_j &= (x_j + \Delta x_j) + i(y_j + \Delta y_j) \\ \bar{c}_j &= (x_j - \Delta x_j) + i(y_j - \Delta y_j), \end{aligned} \tag{7}$$

where  $\Delta x_j$  and  $\Delta y_j$  are  $CP$ -violating parameters.

We generate 100 signal-only datasets from the results of this analysis. The sample size is Poisson-distributed with an expected value of 4585, which is the signal yield obtained in the fit to the *BABAR* data [15]. We then consider each of the 4950 possible pairwise combinations of these samples as datasets from two different experiments. For the purpose of this example we focus on one of the parameters of interest of the *BABAR* analysis: the  $CP$  violating parameter  $\Delta x$  of the  $\rho^0(770)K^\pm$  contribution,  $\Delta x_\rho$ . The value used to generate events is that measured by *BABAR*, namely  $-0.160 \pm 0.049 \pm 0.024^{+0.094}_{-0.013}$ . After fitting each of the individual samples with the model used for event generation, we obtain combined results for  $\Delta x_\rho$  by three different methods:

1. a naïve average of the individual results for  $\Delta x_\rho$ , taking into account the parabolic uncertainties of the fits to the individual samples;
2. a naïve average, taking into account the asymmetric uncertainties of the individual fits<sup>4</sup>, following the prescription from Ref. [8];
3. a joint fit to the two samples in the **JFIT** framework, following the method proposed in this paper, using the full likelihood function.

The results obtained from methods 1 and 2 above are found to be equivalent: in the present case, with a very few exceptions, the effect of the asymmetric nature of the likelihood is negligible comparing to the statistical uncertainty. The results obtained from methods 2 and 3 are compared in Fig. 4, and show a much larger difference. The distribution of the difference between results obtained by the two methods has a full width at half maximum of approximately 10% of the typical statistical uncertainty on  $\Delta x_\rho$  in fits to individual datasets (which is 0.03).

We perform likelihood scans as a function of  $\Delta x_\rho$  for several individual datasets and their corresponding joint fits, i.e., we fix  $\Delta x_\rho$  to several consecutive values, for each of which the fit is repeated. The other parameters are free to vary as in the nominal fit. We compare the sum of likelihood functions obtained in this way from two datasets to the likelihood scan obtained with the corresponding **JFIT**; this is shown in Fig. 5. This comparison illustrates the fact that even if two experiments provide their full likelihood dependences on certain parameters, obtaining a combined result by summing up these functions is not equivalent to performing a joint fit. Indeed, values of nuisance parameters, for which combined results are not desired, generally differ between the joint fit and the individual fits. Note that in the particular case shown in Fig. 5 the minimum obtained from the joint fit is closer to the generated value of  $-0.16$  and has a smaller uncertainty than the other combination. Also, the joint-fit result is smaller than both minima obtained from the individual fits. This indicates that even in well-behaved cases and even if the combination is performed by summing likelihood functions, neglecting correlations may result in biases.

---

<sup>4</sup> The asymmetric uncertainties are obtained from the MINOS routine of the MINUIT package.

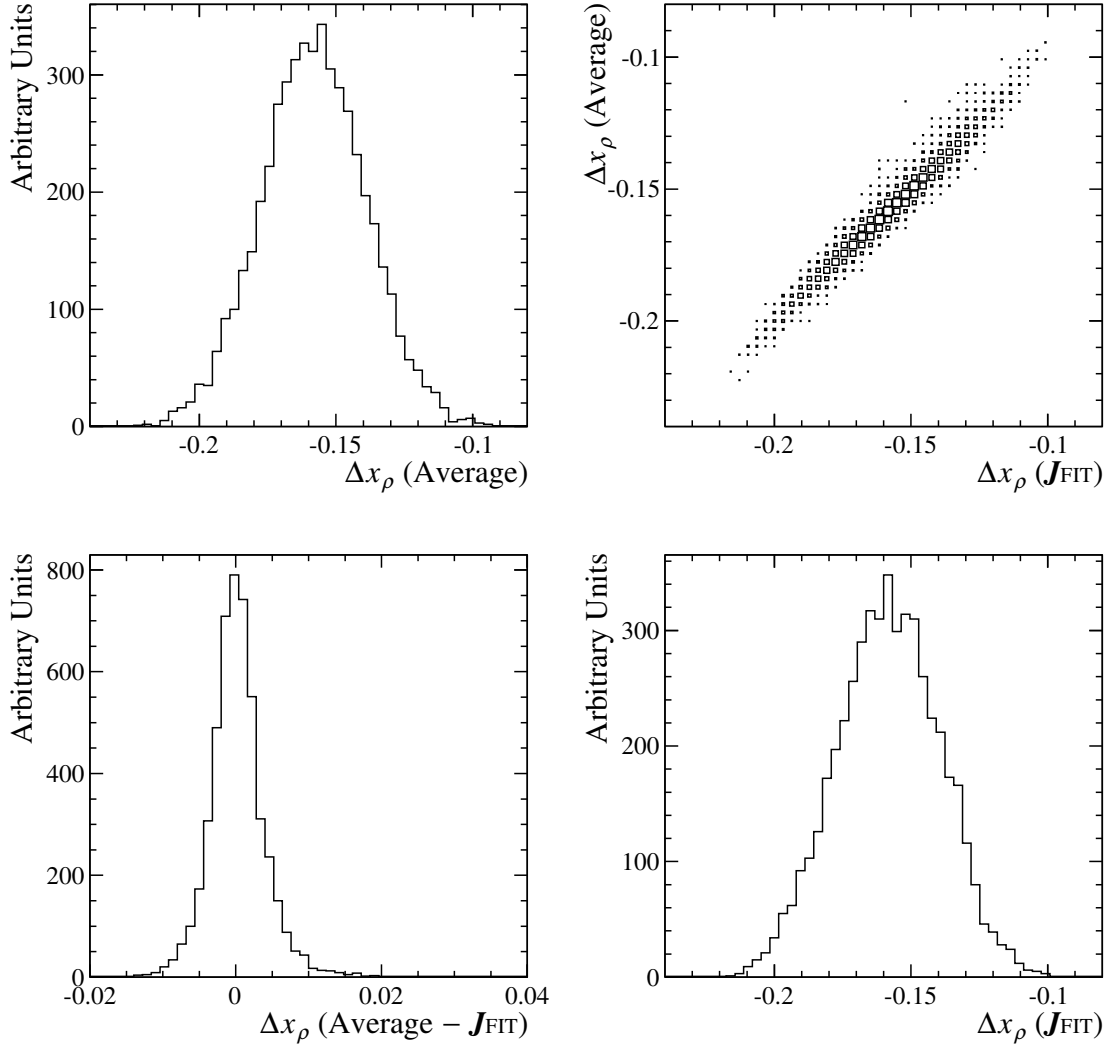


Figure 4: Distributions of  $\Delta x_\rho$  results obtained by (top left) naïve averages taking into account the asymmetric uncertainties of the individual fits to two samples, (bottom right)  $\mathbf{J}_{FIT}$ s (top right) the former versus the latter and (bottom left) the difference between the two.

To evaluate potential bias induced by naïve averaging, the distance of combined results to the generated value was compared between naïve averages, taking into account the asymmetric uncertainty, and  $\mathbf{J}_{FIT}$ s. This study shows that results obtained by joint fits are more often closer to the generated value than those obtained by naïve averages due to the fact that they fully account for the correlations between the fit variables. Figure 6 shows the distance-from-generated-value of  $\Delta x_\rho$  results obtained by naïve averaging versus the corresponding quantity for  $\mathbf{J}_{FIT}$ s, and the distribution of the difference between the former and the latter. To clarify the presence of the non-Gaussian tails in this distribution, it is overlaid with a Gaussian fitted to its central region  $[-0.005, 0.005]$ ; numbers of positive and negative entries in the distribution, excluding the ranges corresponding to one, two and three standard deviations of the Gaussian are given in Table 3. Comparison of the statistical

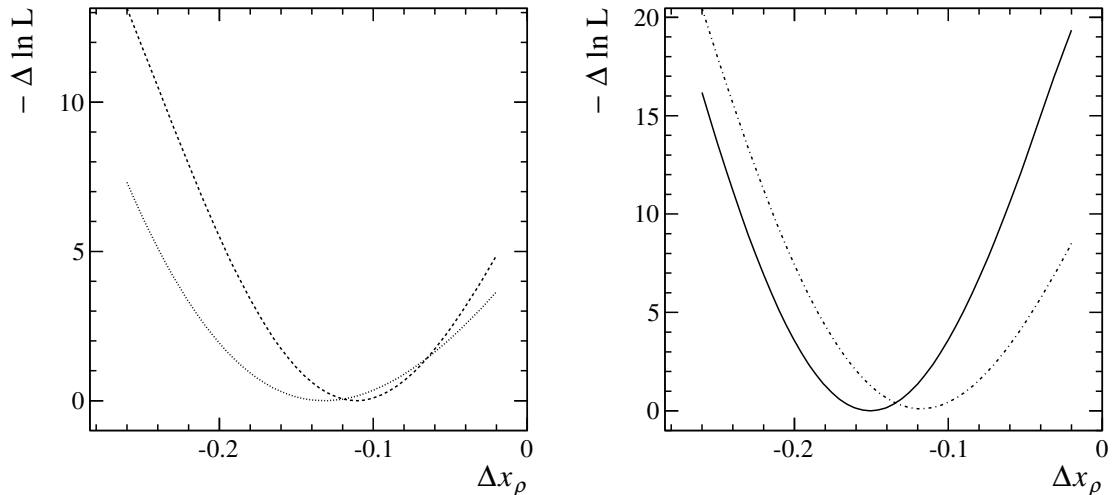


Figure 5: Left: Likelihood scans, showing  $-\Delta \ln \mathcal{L} \equiv (\ln \mathcal{L})_{\min} - \ln \mathcal{L}$  as a function of  $\Delta x_\rho$  in two different datasets. Right: the sum of these likelihoods (dashed-dotted curve), compared to the likelihood obtained from a  $\mathbf{JFIT}$  to the two samples (solid curve). It should be noted that the result obtained from the joint fit is smaller than both of those from the individual fits and is closer to the true value ( $-0.16$ ). In addition it has a smaller uncertainty than the simple average.

Table 3: The distribution on the right hand side of Fig. 6 illustrates the fact that results obtained by joint fits are more often closer to the generated value than these obtained by naïve averages. Here are given the numbers of positive and negative entries in the distribution, excluding the ranges corresponding to one, two and three standard deviations ( $\sigma$ ) of the overlaid Gaussian.

Excluded region ( $\sigma$ )	Number of positive entries	Number of negative entries
3	126	71
2	330	266
1	918	911

uncertainties obtained from naïve averages and  $\mathbf{JFIT}$ s are shown in Fig. 7. In 88% of the cases the uncertainty obtained from a joint fit is smaller.

We stress that the example given here is not extreme: likelihoods are nearly Gaussian and are rather well behaved, as can be seen in Fig. 5. In cases where the likelihood presents strong non-linear features, such as asymmetric functions that cannot be well described by a bifurcated Gaussian, or if it has multiple minima, the difference between naïve averaging and joint fitting could be much larger. In practice, multiple solutions appear in nearly all the Dalitz-plot analyses performed by the  $B$  factories; they represent one of the major difficulties in these analyses. Clearly, a joint fit allows to resolve better the global minimum from the mirror solutions.

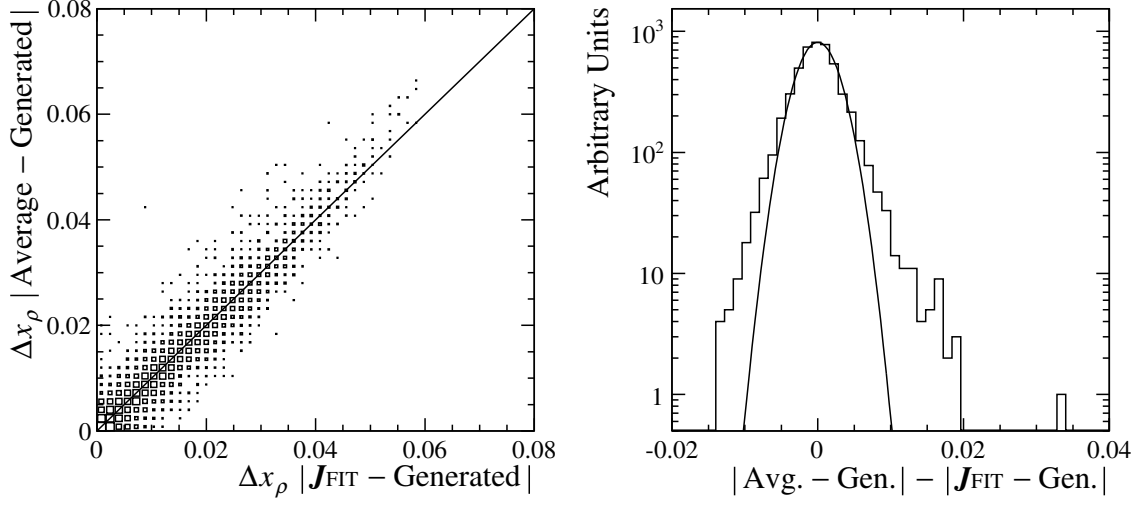


Figure 6: Left: distance from the generated value ( $-0.16$ ) of  $\Delta x_\rho$  results obtained by naïve averaging versus that corresponding to  $\mathbf{J}_{\text{FIT}}$ s. Right: distribution of the difference between the former and the latter. The solid smooth curve is a Gaussian fitted to the central region of the distribution  $[-0.005, 0.005]$ , to clarify the presence of the non-Gaussian tails.

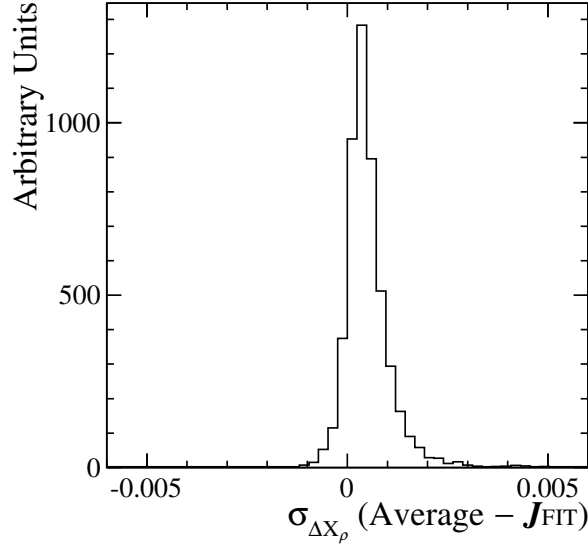


Figure 7: The difference between uncertainties obtained in naïve averages, and those from  $\mathbf{J}_{\text{FIT}}$ s. In the former, the average between the positive and negative asymmetric uncertainties is used. In 88% of the cases joint fits yield improved sensitivity to  $\Delta x_\rho$ .



## References

- [1] The web page of the RooFit project: <http://roofit.sourceforge.net>
- [2] The web page of the ROOT project: <http://root.cern.ch>
- [3] F. James, 1972 MINUIT - Function Minimization and Error Analysis CERN Program Library Long Writeup D506; C++ translation of MINUIT <https://root.cern.ch/root/html/TMinuit.html>
- [4] The web page of the Laura++ project: <http://laura.hepforge.org>
- [5] The Doxygen documentation for the `LauSimFitMaster` class in Laura++: <http://laura.hepforge.org/doc/doxygen/v3r0p1/classLauSimFitMaster.html>
- [6] The Doxygen documentation for the `LauSimFitSlave` class in Laura++: <http://laura.hepforge.org/doc/doxygen/v3r0p1/classLauSimFitSlave.html>
- [7] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. D **91**, 092002 (2015).
- [8] K. A. Olive *et al.* [Particle Data Group Collaboration], Chin. Phys. C **38**, 090001 (2014).
- [9] M.J. Oreglia, Ph.D Thesis, SLAC-236(1980), Appendix D; J.E. Gaiser, Ph.D Thesis, SLAC-255(1982), Appendix F; T. Skwarnicki, Ph.D Thesis, DESY F31-86-02(1986), Appendix E.
- [10] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **716**, 1 (2012).
- [11] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Lett. B **716**, 30 (2012).
- [12] Ed. A.J. Bevan, B. Golob, Th. Mannel, S. Prell, and B.D. Yabsley, Eur. Phys. J. C **74**, 3026 (2014).
- [13] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. D **87**, 072004 (2013).
- [14] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. D **90**, 072003 (2014).
- [15] B. Aubert *et al.* (BABAR Collaboration), Phys. Rev. D **78**, 012004 (2008).
- [16] A. Garmash *et al.* (Belle Collaboration), Phys. Rev. Lett. **96**, 251803 (2006).